

Twitter Sentiment Analysis

Krunal Dhardev¹, Dr. Kamalraj R²

¹Student, ²Associate Professor,

^{1,2}School of CS and IT, Jain University, Bangalore, Karnataka, India

ABSTRACT

Microblogging today has gotten an acclaimed specific instrument among Internet clients. Endless clients share assessments on various bits of life dependably. Accordingly, microblogging districts are rich wellsprings of information for assessment mining and tendency assessment. Since microblogging has shown up by and large lately, there several investigation works that were given to this point. In our paper, we base on using Twitter, the most notable microblogging stage, for the task of feeling examination. We advise the most ideal approach to thus accumulate a corpus for assessment and evaluation mining purposes. We play out a semantic assessment of the amassed corpus and clarify found wonders. Utilizing the corpus, we build up an end classifier, that can pick positive, negative, and honest evaluations for an annual. Test assessments show that our proposed strategies are convincing and act in a way that is better than actually proposed procedures. In our appraisal, we worked with English, in any case, the proposed procedure can be utilized with some other language.

KEYWORDS: social media; Sentiment Analysis; Twitter

How to cite this paper: Krunal Dhardev | Dr. Kamalraj R "Twitter Sentiment Analysis" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-4, June 2021, pp.758-761, URL: www.ijtsrd.com/papers/ijtsrd42385.pdf



Copyright © 2021 by author (s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



1. INTRODUCTION

- Assessment and nostalgic mining are significant exploration territories in light of the fact that because of the gigantic number of day-by-day posts on interpersonal organizations, separating individuals' feelings is a difficult errand. Around 90% of the present information has been given during the most recent two years and getting knowledge into this enormous scope of information isn't unimportant
- The nostalgic examination has various applications for different regions for example in associations to get reactions for things by which associations can get comfortable with customer's information and reviews by means of online media.
- Appraisal and nostalgic mining have been a lot of packed in this reference and each different technique and investigation fields have been discussed. There are likewise a few works that have been done on Facebook nostalgic investigation anyway in this paper we generally canter around the Twitter wistful examination.
- For bigger writings, one arrangement could be to comprehend the content, sum up it, and offer load to it whether it is positive, negative, or impartial. Two essential ways to deal with remove text rundown are extractive and abstractive techniques. In the extractive technique, words and word phrases are removed from the first content to produce a rundown. In an abstractive strategy, attempts to get familiar with an inward language portrayal and afterward creates an outline that is more like the synopsis done by people.
- Text understanding is a huge issue to tackle. Some AI procedures, including different directed and unaided calculations, are being used. There are various ways to deal with produce an outline. One methodology could be to rank the significance of sentences inside the content and afterward create a rundown for the content dependent on the significant numbers. There is another methodology called start to finish generative models. In a few an area like picture acknowledgment, discourse acknowledgment, language interpretation, and question-replying, the start to finish strategy performs better.
- A couple of works have used cosmology to fathom the substance. At the articulation level, the contemplative examination system should have the alternative to see the furthest point of the articulation which is discussed by Wilson, et.al. Tree part and feature-based model have been applied for insightful examination in Twitter by Agarwal and et.al. SemEval-2017 also shows the seven years of nostalgic examination in Twitter tasks. Since tweets on Twitter is a specific version of a common book there are a couple of works that address this issue like the work for short easy-going compositions. The nostalgic examination has numerous applications in news.
- In this paper, we will talk about interpersonal organization examination and its significance, at that point we talk about Twitter as a rich asset for nostalgic investigation. In the accompanying areas, we show the significant level unique of our execution. We will show a few questions on various points and show the extremity of tweets.

- Notion Analysis is a strategy generally utilized in text mining. Opinion Analysis is an NLP and data extraction task that expects to acquire essayists felling communicated in certain or negative remarks, questions, and demands by breaking down an enormous number of archives.
- As a rule, notion investigation expects to decide the demeanour of a speaker or essayist as for some theme or the general usefulness of the report. Fundamentally, supposition Analysis is the undertaking of recognizing whether the assessment communicated in content is positive or negative.
- We utilize a dataset shaped by gathered messages from Twitter. Twitter contains an enormous number of extremely short messages made by the clients of this microblogging stage. The substance of the messages change from individual contemplations to public proclamations.
- Ideological groups might be intrigued to know whether individuals support their program or not. Social associations may ask individuals' feelings on current discussions. This data can be acquired from microblogging administrations, as their clients post each day what they like/hate, and their suppositions on numerous parts of their life.

- In our paper, we concentrate on how microblogging can be utilized for notion examination purposes. We tell the best way to utilize Twitter as a corpus for estimation examination and assessment mining.

2. DATA CAPTURING AND PROCESSING

2.1. Data Capturing Process:

Twitter streaming API is utilized to catch the information. Twitter real time API assists with making associations between PC projects and web administrations. For getting to Twitter streaming API we need four keys called API Key, API Secret, Access Token, and Access Token Secret. Steps to recover four keys

- Create a Twitter account
- Open page <https://apps.twitter.com/> and login with twitter credentials
- Try creating a new app
- Fill the form and 'Create twitter new application'
- Retrieve API keys and API secret
- Retrieve access token and Access token secret.

Once all four keys are retrieved, I have used a python library called Tweepy to download the tweets. Tweepy is connected to twitter streaming API to retrieve particular product latest post by default I set as 25 but we can increase the no of post we want.

Here is the example code used to retrieve resent 25 tweets from Twitter.

```
In [4]: # Extract 50 tweets from the twitter user
str1="#COVIDIOTS";
if(str1[0]=="#"):
    posts=api.search(q=str1,count=25,lang="en",tweet_mode="extended")
else:
    posts = api.user_timeline(screen_name=str1, count = 25, lang = "en", tweet_mode="extended")

# Print the last 10 tweets
print("Show the 5 recent tweets:\n")
i=1
for tweet in posts[:5]:
    print(str(i) + ' ' + tweet.full_text + '\n')
    i=i+1
```

Show the 5 recent tweets:

- 1) @ember_november I'm not going to complain about the inconsistencies because at least we finally have meaningful restriction s. Except the malls - they left a gaping hole for all those #COVIDIOTS
- 2) @truth_vaccine @CMANN66 @ctwinnipeg @jeffkeelectv And just like that the covidiot went away. #covidiot
- 3) RT @spmayurbhanj: His name is Paramananda Ghosh. He sells vegetable near Murgabadi Golei. He wears Mask properly. He wears face shield for...
- 4) Is there any political party dare to prosecute this regime?#COVIDIOTS <https://t.co/5rbEBtAwrs>
- 5) @JamesStVincent @sebastianbach True that! Fully vaccinated here...but still looking forward to the 5G n Bluetooth that Q is promising. 😊

#COVIDIOTS

Figure 1. Example Code Tweet Retrieval

2.2. Data Processing:

Tweets are caught in a Data outline. Tweets are characterized into 3 classes. Positive, negative, and neutral as noticed previously. Table 1 show some illustration of how tweets would be arranged in their classifications.

Table 1 Tweets Classification Example

```
In [30]: # Create a function to compute negative (-1), neutral (0) and positive (+1) analysis
def getAnalysis(score):
    if score < 0:
        return 'Negative'
    elif score == 0:
        return 'Neutral'
    else:
        return 'Positive'

sent_df['Analysis'] = sent_df['sentiment_value'].apply(getAnalysis)
# Show the dataframe
sent_df.head()
```

```
Out[30]:
```

	Tweets	sentiment_value	word_count	index	Analysis
0	_november I'm not going to complain about the ...	0.0	17	1	Neutral
1	Except the malls - they left a gaping hole for...	2.0	23	2	Positive
2	covidiot His name is Paramananda Ghosh.	0.0	6	3	Neutral
3	He sells vegetable near Murgabadi Golei.	0.0	6	4	Neutral
4	He wears Mask properly.	0.0	4	5	Neutral

2.2.1. Pre-Processing :

Each tweet goes through the accompanying pre-handling steps:

1. Remove unnecessary data :
 - Removing word that contains @
 - Removing '#' hash tag
 - Removing hyperlink (https:)
 - Removing RT (retweet)
 - Remove \n
 - Remove :
 - Remove both the leading and the trailing characters
 - Removes empty strings, because they are considered in Python as False

sub() strategy from python's standard articulation class was utilized to substitute URLs, usernames, void areas, hashtags with the pertinent qualities as clarified previously. The strip() technique from string class is utilized to strip the leftover expressions of any accentuation.

2.2.2. Obtaining Stop Words List:

A stop word is an ordinarily utilized word, (for example, "the", "a", "an", "in") that an internet searcher has been customized to disregard, both when ordering sections for looking and while recovering them as the consequence of an inquiry question.

We would not need these words to occupy room in our data set, or occupying the important preparing time. For this, we can eliminate them effectively, by putting away a rundown of words that you consider to stop words. NLTK(Natural Language Toolkit) in python has a rundown of stopwords put away in 16 unique dialects.

2.2.3. Stemming:

Stemming is the route toward perceiving induced words and consigning a word to all of the decided words. This will diminish the size of the rundown reports. The basic tweet which is in the string configuration is changed over into a python once-over of substrings which can be used to procure all of the words and highlight in the tweet. The NLTK Tokenizer Package is used thus. The hidden string is decoded to utf8 to make an effort not to manage encoded strings. NLTK library as of now gives an execution of the Porter stemmer computation in the nltk.stem.porter module. The tokenized string is used as a commitment to the Porter stemmer.

Methodology:**1. tf-idf:**

tf-idf represents Term recurrence opposite report recurrence. The tf-idf weight is a weight frequently utilized in data recovery and text mining. Varieties of the tf-idf weighting plan are regularly utilized via web crawlers in scoring and positioning a report's pertinence given a question. This weight is a factual measure used to assess how significant a word is to a report in an assortment or corpus. The significance builds relatively to the occasions a word shows up in the archive yet is balanced by the recurrence of the word in the corpus (informational index).

tf-idf is a weighting plan that appoints each term in a record a weight dependent on its term recurrence (tf) and backwards report recurrence (idf). The terms with higher weight scores are viewed as more significant.

2. Bag of words :

The pack of-words model is an improving on portrayal utilized in normal language handling and data recovery (IR).

In this model, a content (like a sentence or a record) is addressed as the pack (multiset) of its words, dismissing syntax and even word request yet keeping assortment. The pack of-words model has additionally been utilized for PC vision.

The sack of-words model is usually utilized in strategies for record grouping where the (recurrence of) event of each word is utilized as an element for preparing a classifier.

OBJECTIVES

Sentiment analysis over Twitter offer organisations a fast and effective way to monitor the publics' feelings towards their brand, business, directors, etc. A wide range of features and methods for training sentiment classifiers for Twitter datasets have been researched in recent years with varying results.

Business

Politics

Public Actions

CONCLUSION:

- In this specific paper, we inspected the meaning of casual local area examination and its applications in different areas. We focused on Twitter and have executed the python program to do the insightful assessment. We showed the results on different step-by-step subjects. We comprehended that the impartial speculations are in a general sense high which shows there is a need to improve Twitter thought examination.

- Twitter conclusion examination is created to research the public's perspectives towards a tweet/hashtag. Info is given i.e., either the username or a hashtag. At that point, the tweet is recovered from twitter data that goes through highlight extraction. Partner in Nursing prudent element vector is framed by doing highlight extraction in 2 stages when right pre-handling. inside the beginning, the Twitter-explicit alternatives territory unit removed and added to the component vector. From that point forward, these alternatives region unit detached from tweets, and again highlight extraction is done as though it's done on conventional content. These alternatives are extra to the element vector. Characterization precision of the component vector is tried exploitation Naïve Thomas Bayes classifier. Partner in Nursing precision of 78. 38 it had been reached.

Future scope:

It can implement on social media like Facebook and Instagram, and also try to capture past post.one more thing that I can implement in future. That thing is if any person post a negative tweets we can try to change diplomatic sentence. For diplomatic sentence we will use neural networks.

Acknowledgement:

Foremost, I would like to thank Dr.M. N Nachappa and Pro. Kamalraj R for the continuous support of my MCA research paper, and for their knowledge, patience, and motivation throughout the year. Your guidance helped me in research and writing this thesis, their convenient course, total co-activity, and moment perception have made my work productive.

References

- [1] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of twitter data
- [2] Guerra, P., Veloso, A., Meira Jr, W., Almeida, V.: From bias to opinion: A transfer-learning approach to real-time sentiment analysis.
- [3] Guerra, P., Veloso, A., Meira Jr, W., Almeida, V.: From bias to opinion: A transfer-learning approach to real-time sentiment analysis.
- [4] Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: The good the bad and the omg! In: Proceedings of the ICWSM.s
- [5] David Vilares, Yeraí Doval, Miguel A. Alonso, and Carlos Gomez-Rodriguez.
- [6] Adil Moujahid, An Introduction to Text Mining using Twitter Streaming API and Python
- [7] Sunil Ray, September 2017, Understanding Support Vector Machine Algorithm from Examples

